



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/091,472	03/07/2002	Jae Shin Yu	HI-0074	9044

34610 7590 05/16/2006

FLESHNER & KIM, LLP
P.O. BOX 221200
CHANTILLY, VA 20153

EXAMINER

TRAN, QUOC A

ART UNIT	PAPER NUMBER
----------	--------------

2176

DATE MAILED: 05/16/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/091,472	Applicant(s) YU ET AL.	
	Examiner Quoc A. Tran	Art Unit 2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 21 February 2006.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-23 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-23 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This action is responsive to Amendment filed 2/21/2006, with acknowledgement of original filing date of 03/07/2004, which was benefited from foreign priority No. 12184/2001 filed 03/09/2001.
2. Claims 1-23 are currently pending in this application. Claims 21-23 are new and claims 1, 13 and 17 are independent claims.
3. 35 U.S.C. 101 rejections of claims 1-20 has been withdrawn.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. **Claims 1-23** are rejected under 35 U.S.C. 103(a) as being unpatentable by Gibbon et al. US006714909B1 – filed 11/21/2000 (hereinafter Gibbon ‘909), in view of Nelson et al. US006243713B1 – filed 08/24/1998 (hereinafter Nelson ‘713).

In regard to independent claim 1, extracting a plurality of text areas from a video stream (Gibbon ‘909 at col. 2, lines 1-30, discloses ability to segment multimedia data, such as news broadcasts, into retrievable units that are directly related to what users perceive as meaningful, such as separating a multimedia data stream into audio, visual and text components, segmenting the audio, visual and text components based on semantic differences),

calculating importance measures according to weights for each of the extracted text areas (Gibbon '909 at col. 8, line 45 through col. 14, line 40, also see Fig. 13-Fig.17, discloses a mechanism to recover the semantic structure of the data for creating appropriate descriptions of the extracted multimedia content, such as:

- (i) To present the semantic structure to the users,
- (ii) To represent the particular semantics based on the content of the news story,
- (iii) To form the representation for news summary of the day,

Wherein textual and another is combination of text with visual is to automatically construct the representation in a form that is most relevant to the content of the underlying story according to their importance computed as weighted frequency (see Gibbon '909 at col. 8, line 10 through col. 9 through col. 12, line 5 for detail of the calculation steps and formula of the importance computed as weighted frequency) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein calculating importance measures would have been an obvious variant of computed as weighted frequency.

Gibbon '909 does not explicitly teach, **synthesizing the number of text areas into a synthetic key frame**, however (Nelson '713 at col. 6, lines 5-50, discloses compound documents which are separated into constituent multimedia components of different data types, such as text, images, video, audio/voice, and other data types with portion thereof. Preferably these various multimedia components are combined with one or more query operators includes both text and image components, and a number of query operators defining both logical relationships and proximity relationships between the multimedia components) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein synthetic key frame would

have been an obvious variant of separated into constituent multimedia components of different data types and combined with one or more query operators includes both text and image components),

Selecting a number of text areas to be synthesized based upon the importance measures in the order of higher importance, however (Nelson '713 at col. 6, lines 5-50, however (Nelson '713 at col. 6, lines 5-50, discloses compound documents which are separated into constituent multimedia components of different data types, such as text, images, video, audio/voice, and other data types with portion thereof. Preferably these various multimedia components are combined with one or more query operators includes both text and image components, and a number of query operators defining both logical relationships and proximity relationships between the multimedia components) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein synthetic key frame and based upon the importance measures would have been an obvious variant of separated into constituent multimedia components of different data types and combined with one or more query operators includes both text and image components and logical relationships and proximity relationships between the multimedia components, to a person of ordinary skill in the art at the time the invention was made .

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified Gibbon '909, discloses a method of extracting a plurality of text areas from a video stream and calculating importance measures according to weights for each of the extracted text areas, to include a means of synthesizing the text areas to be synthesized into the key frame and selecting the number of text areas to be synthesized based

Art Unit: 2176

upon the importance measures in the order of higher importance of Nelson '713. One of the ordinary skills in the art would have been motivated to perform such a modification to provide a desirable system that retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video (as taught by Nelson '713 at col. 2, lines 10-20).

In regard to independent claims 13 and 17, incorporate substantially similar subject matter as cited in claim 1 above, and are similarly rejected along the same rationale.

In regard to dependent claim 2, wherein the text areas are extracted according to certain intervals of the video stream (Gibbon '909 at col. 14, lines 15-35, discloses the steps wherein during playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).

In regard to dependent claim 3, wherein a synthetic key frame is generated for each of the certain intervals of the video stream (Gibbon '909 at col. 13, lines 15-35 discloses the steps during playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).

In regard to dependent claim 4, wherein the certain intervals of the video stream are discriminated by scenes as logical units of a video (Gibbon '909 at col. 10, line 15 through col. 11, line 40, also see Fig. 12, discloses a stream of detected audio events where A stands for anchor's speech, D stands for detailed reporting (from non-anchor people), and C stands for commercials. The center timeline in FIG. 12 shows the segments of text obtained from the text

Art Unit: 2176

event segmentation (unit 405) using marker A where the duration of each segment does not include commercials) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein certain intervals and logical units would have been an obvious variant of timeline the segments of text obtained from the text event segmentation using markers A, C and D, to a person of ordinary skill in the art at the time the invention was made .

In regard to dependent claim 5, wherein the certain intervals of the video stream are discriminated by shots as physical units of a video (Gibbon '909 at col. 10, line 15 through col. 11, line 40, also see Fig. 12, discloses a stream of detected audio events where A stands for anchor's speech, D stands for detailed reporting (from non-anchor people), and C stands for commercials. The center timeline in FIG. 12 shows the segments of text obtained from the text event segmentation (unit 405) using marker A where the duration of each segment does not include commercials) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein certain intervals and logical units would have been an obvious variant of timeline the segments of text obtained from the text event segmentation using markers A, C and D, to a person of ordinary skill in the art at the time the invention was made .

In regard to dependent claim 6, incorporate substantially similar subject matter as cited in claim 1 above and in further view of the following, and is similarly rejected along the same rationale, ...a display duration time of a text (Gibbon '909 at col. 10, line 15 through col. 11, line 40, also see Fig. 12, discloses a center timeline wherein the segments of text obtained from the text event segmentation (unit 405) using marker A, C and D...).

In regard to dependent claim 7, wherein the mean text size in the text area is determined by using a density and size of a histogram for the text area (Gibbon '909 at col.

Art Unit: 2176

13, line 60 through col. 14, line 5, also see Fig. 15, discloses a keyword histogram is first constructed a fixed number of key frames within the boundary are chosen so that they (1) are not within anchor speech segments and (2) yield maximum covered area with respect to the keywords histogram. The peak points marked on the histogram in FIG. 15 indicate the positions of the chosen frames and the shaded area underneath them defines the total area coverage on the histogram by the chosen key frames...).

In regard to dependent claim 8, wherein the display duration time of the text is determined by considering whether a previously extracted text area is identical to a currently extracted text area (Gibbon '909 at col. 10, lines 50-65, discloses the block of text available at this point, the task is to determine how these blocks of text can be merged to form semantically coherent content based on appropriate criteria. Since news introductions are to provide a brief and succinct message about the story, they naturally have a much shorter duration than the detailed news reports. Based on this observation, in step 5060, a headline story segmentation unit 440 initially classifies each block of text as a story candidate or an introduction candidate based on duration. ..., Also Gibbon '909 at col. 12, lines 15-25 discloses blocks formed in this way not only contain enough information for similarity comparison but also have natural breaks of chains of repeated words if true boundaries are present...) Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein determined by considering whether a previously extracted text area is identical to a currently extracted text area would have been an obvious variant of t block of text available at this point and contain enough information for similarity comparison but also have natural breaks of chains

of repeated words if true boundaries are present , to a person of ordinary skill in the art at the time the invention was made .

In regard to dependent claim 9, wherein the weight increases as the size of the text area, the mean text size of the text area or the display duration time of the text increases, (Gibbon '909 at col. 13, lines 30-50, also see Fig. 14, discloses a window that plays back streaming content to a user. It is triggered when users click on a particular item. In this playback window, the upper portion shows the video and the lower portion the text synchronized with the video. During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given... keywords are chosen in step 5080 above, from the story according to their importance computed as weighted frequency).

In regard to dependent claims 10-12, 15-16 and 18 incorporate substantially similar subject matter as cited in claim 1 above, and are similarly rejected along the same rationale. Examiner read the above in the broadest reasonable interpretation to the claim limitation, wherein the certain rule is addition of values obtained by multiplying the weight determining factors with the corresponding weights and wherein the weight determining factors would have been an obvious variant of calculating importance measures according to weights, to a person of ordinary skill in the art at the time the invention was made.

In regard to dependent claim 14, incorporate substantially similar subject matter as cited in claim 6 above, and is similarly rejected under the same rationale.

In regard to dependent claim 19, incorporate substantially similar subject matter as cited in claim 7 above, and is similarly rejected under the same rationale.

In regard to dependent claim 20, incorporate substantially similar subject matter as cited in claim 8 above, and is similarly rejected under the same rationale.

6. **Claims 21-23** are rejected under 35 U.S.C. 103(a) as being unpatentable by Gibbon et al. US006714909B1 – filed 11/21/2000 (hereinafter Gibbon '909), in view of Nelson et al. US006243713B1 – filed 08/24/1998 (hereinafter Nelson '713), and further in view of Maybury et al. US006961954B1 filed 03/02/1998 (hereinafter Maybury).

In regard to dependent claim 21-23, incorporate substantially similar subject matter as cited in claim 1, 13 and 17 above, and further view of the following and are similarly rejected under the same rationale;

wherein the synthetic key frame is used by a browser to search for multimedia information, however (Maybury at col. 16 lines 34-50), discloses The Broadcast News Navigator 200 enables a user to search and browse the meta data files 142 via a computer network. The user may do so through a graphical interface using a Web browser such as Netscape, Microsoft Explorer, or NCSA Mosaic, wherein the graphical interface is provided by creating Hypertext Markup Language (HTML) page files and/or Java applets that access the meta data 142 in a manner that is well known in the art.

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified Gibbon '909 and Nelson '713, to include a means of synthesizing the synthetic key frame is used by a browser to search for multimedia information of Maubury's teaching. One of the ordinary skills in the art would have been motivated to

perform such a modification to provide a desirable system that retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video (as taught by Nelson '713 at col. 2, lines 10-20).

Response to Argument

7. Applicant's arguments filed 02/21/2006 have been fully considered but they are not persuasive. The reason is set forth in the current Office Action cited below and further view of the following:

The examiner respectfully notes that Applicant has added new claims 21-23. To address these amendments, the Examiner introduces the Maybury reference (see rejection above for detail).

8. Brief summary of prior art of records:

Gibbons discloses the method and system for use with the broadcasts news program browser, that provides users with the ability to retrieve information from multimedia events, such as broadcast news programs, in a semantically meaningful way at different levels of abstraction, wherein likelihood method and threshold based method is used (i.e. Gaussian Mixture Model (GMM) is employed to model news and commercial classes, individually. A GMM model consists of a set of weighted Gaussian, wherein the mean vector and covariance matrix applied).

Nelson discloses the method and system for indexing multimedia document and multimedia query that includes text, image, video, audio and other (i.e. extract different type of media, and unified multimedia index, search and score then producing results).

Maybury discloses the method and system for use with Broadcast News Navigation (BNN), that enables a user to search and browse the meta data files via a computer network. The user may do so through a graphical interface using a Web browser such as Netscape, Microsoft Explorer.

5. Response to Arguments:

Beginning on page 8 of the Remarks (hereinafter the remarks), Applicant argues the following issues, which are accordingly addressed below.

Applicant's arguments, on pages 9-13 of the remarks that Gibbon in combination with Nelson do not teach:

(1) Gibbon fails to teach Calculating importance measures according to weights for each of the extracted text areas;

(2) Nelson does not teach (1) and selecting areas to be synthesized and synthesizing a key frame (the same arguments are substantially repeated for independent claims 7, and 13 pending).

The examiner respectfully disagrees. **In response to (1)**, The examiner respectfully notes that Using the broadest interpretation, Gibson at col. 8, line 45 through col. 14, line 40, also see Fig. 13-Fig.17, discloses a mechanism to recover the semantic structure of the data for creating appropriate descriptions of the extracted multimedia content, such as:

- (i) To present the semantic structure to the users,
- (ii) To represent the particular semantics based on the content of the news story,
- (iii) To form the representation for news summary of the day,

Wherein textual and another is combination of text with visual is to automatically construct the representation in a form that is most relevant to the content of the underlying story according to their importance computed as weighted frequency (see Gibbon '909 at col. 8, line 10 through col. 9 through col. 12, line 5 for detail of the calculation steps and formula of the importance computed as weighted frequency) and

for further more supports, (see Gibbon '909 at col. 14 line 65 through co. 15 line 25), disclose the semantically coherent text blocks based on a set of topic category models; generating a multimedia description of the multimedia event based on the identified target speaker, the semantically coherent text blocks, the identified topic, and the generated summary; and extracting audio features from the audio component of the multimedia data stream, the audio features being at least one of frame-level and clip level features, wherein the frame level features in three subbands are at least one of volume, zero crossing rate, pitch period, frequency centroid, frequency bandwidth, and energy ratios, and (see Gibbon '909 at col. 8, line 10 through col. 9 through col. 12, line 5 for detail of the calculation steps and formula of the importance computed as weighted frequency).

Using the broadest interpretation of the claimed invention, One of the of ordinary skill in the art at the time the invention was made would have been appreciated that Gibbon is employed A GMM model consists of a set of weighted Gaussian, wherein the mean vector and covariance matrix for determining the semantically coherent text blocks, the identified topic, and the generated summary; and extracting audio features from the audio component of the multimedia data stream, the audio features being at least one of frame-level and clip level features, wherein the frame level features in three subbands are at least one of volume, zero crossing rate, pitch period, frequency centroid, frequency bandwidth, and energy ratios.

The examiner respectfully notes that, using the broadest interpretation, Gibson art structure is capable of performing the intended use, and then it meets the claim.

In response to (2), Nelson does not teach (1) (see above for substantially similar arguments and response) and selecting areas to be synthesized and synthesizing a key frame (see above rejection for detail) and further more of the following,

The examiner respectfully notes that Using the broadest interpretation, Gibson taught at col. 2 lines 5-30, the ability to segment multimedia data, such as news broadcasts, into retrievable units that are directly related to what users perceive as meaningful and identifying at least one target speaker using the audio and visual components, also Gibbon taught at col. 12, lines 10-30, the steps of

(1) adaptive granularity that is directly related to the content is achieved,

(2) the hypothesized boundaries are more natural than those obtained using a fixed window, commonly adopted in a conventional discourse segmentation method,

(3) blocks formed in this way not only contain enough information for similarity comparison but also have natural breaks of chains of repeated words if true boundaries are present,

(4) the original task of discourse segmentation is achieved by boundary verification, and

(5) once a boundary is verified, its location is far more precise than what conventional discourse segmentation algorithms can achieve. This integrated multimodal analysis provides an excellent starting point for the similarity analysis and boundary detection;

but Gibbon does not explicitly teach, selecting areas to be synthesized and synthesizing a key frame, however (see above rejection for detail) and further more of the following, Nelson discloses the method and system for indexing multimedia document and multimedia query that includes text, image, video, audio and other (i.e. extract different type of media, and unified multimedia index, search and score then producing results), using the broadest interpretation, the Examiner reads selecting areas to be synthesized and synthesizing a key frame would have been an obvious variant of the above to one of the of ordinary skill in the art at the time the invention was made, since synthesized is known as to combine digital pulse to result in a new combination, which is Gibbon and Nelson art structure is capable of performing the intended use, and then it meets the claim.

Therefor the Examiner respectfully maintains the rejection of independent claims 1, 13, and 17 for at least the reason cited above at this time.

Furthermore Applicant's arguments, on page 11 of the remarks that Gibbon in combination with Nelson do not teach claim 6 and 9 particularly:

(3) the weights are determined in proportion to the size of the text area, the mean text size of the text area and the display duration time of a text.

(4) the weight increases as the size of the text area, the mean text size in the text area or the display duration time of the text increases.

The examiner respectfully disagrees. **In response to (3) and (4)** (see the rejection above) and further view of the following,

The examiner respectfully notes that Using the broadest interpretation, The examiner respectfully notes that Using the broadest interpretation, Gibson taught at col. 2 lines 5-30, the ability to segment multimedia data, such as news broadcasts, into retrievable units that are directly related to what users perceive as meaningful and identifying at least one target speaker using the audio and visual components, also Gibbon taught at col. 12, lines 10-30, the steps of

(1) adaptive granularity that is directly related to the content is achieved,

(2) the hypothesized boundaries are more natural than those obtained using a fixed window, commonly adopted in a conventional discourse segmentation method,

(3) blocks formed in this way not only contain enough information for similarity comparison but also have natural breaks of chains of repeated words if true boundaries are present,

(4) the original task of discourse segmentation is achieved by boundary verification, and

(5) once a boundary is verified, its location is far more precise than what conventional discourse segmentation algorithms can achieve. This integrated multimodal analysis provides an excellent starting point for the similarity analysis and boundary detection; which Gibbon art structure is capable of performing the intended use, and then it meets the claim.

Therefor the Examiner respectfully maintains the rejection of dependent claims 6 and 9 for at least the reason cited above at this time.

Conclusion

6. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quoc A. Tran whose telephone number is (571) 272-4103. The examiner can normally be reached on Monday through Friday from 9 AM to 5 PM EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Herndon R. Heather can be reached on (571) -272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2176

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Quoc A. Tran
Patent Examiner
Technology Center 2176
May 12, 2006

William L. Bashore
WILLIAM BASHORE
PRIMARY EXAMINER
5/14/2006